Request Form for Translation

Translation Branch
The world of foreign prior art to you.

U. S. Serial No. : 09/901878

Requester's Name: Jon Chang
Phone No. : 305-8439
Fax No. :
Office Location: CPK1 - 4D46
Art Unit/Org. : 2623
Group Director: Joseph Rolla
Is this for Board of Patent Appeals? No

Date of Request: 11/7/02
Date Needed By: 11/21/02
(Please do not write ASAP-indicate a specific date)

PTO 2003-567
S.T.I.C. Translations Branch

Phone:	308-0881
Fax:	308-0989
Location:	Crystal Plaza 3/4 Room 2C01

SPE Signature Required for RUSH:

Document Identification (Select One):
(Note: Please attach a complete, legible copy of the document to be translated to this form)

1. ✓ **Patent** — Document No. 58-56189; Language Japanese; Country Code JP; Publication Date 4/2/83; No. of Pages _____ *(filled by STIC)*

2. ____ **Article** — Author _____; Language _____; Country _____

3. ____ **Other** — Type of Document _____; Country _____; Language _____

RECEIVED 2002 NOV 12 PM 2:46 TRANSLATIONS DIVISION USPTO SCIENTIFIC LIBRARY

Document Delivery (Select Preference):
____ Delivery to nearest EIC/Office Date: 11.21.02 *(STIC Only)*
____ Call for Pick-up ✓ Date: _____ *(STIC Only)*

To assist us in providing the most cost effective service, please answer these questions:

Will you accept an English Language Equivalent? No (Yes/No)

Will you accept an English abstract? No (Yes/No)

Would you like a consultation with a translator to review the document prior to having a complete written translation? No (Yes/No)

Check here if Machine Translation is not acceptable: (It is the default for Japanese Patents, '93 and onwards with avg 5 day turnaround after receipt) (Yes/No)

STIC USE ONLY KLI

Copy/Search
Processor: KLJ
Date assigned: 11/12
Date filled: _____
Equivalent found: N/A (Yes/No)

Doc. No.: _____
Country: _____

Remarks: Copy of Trans E-Mail 11.21.02

Translation
Date logged in: 11.12.02
PTO estimated words: 2937
Number of pages: 9
In-House Translation Available: _____

In-House:
Translator: _____
Assigned: _____
Returned: _____

Contractor:
Name: MC
Priority: S
Sent: 11.13.02
Returned: 11.21.02

PTO 03-567

Japanese Kokai Patent Application
No. Sho 58[1983]-56189

POST-CHARACTER-RECOGNITION PROCESSING SYSTEM

Hideaki Sugahara and Eiichiro Yamamoto

UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. NOVEMBER 2002
TRANSLATED BY THE RALPH MCELROY TRANSLATION COMPANY

JAPANESE PATENT OFFICE
PATENT JOURNAL (A)
KOKAI PATENT APPLICATION NO. SHO 58[1983]-56189

Int. Cl.[3]:	G 06 K 9/62 9/68
Sequence Nos. for Office Use:	7323-5B
Filing No.:	Sho 56[1981]-155629
Filing Date:	September 30, 1981
Publication Date:	April 2, 1983
No. of Inventions:	1 (Total of 5 pages)
Examination Request:	Not filed

POST-CHARACTER-RECOGNITION PROCESSING SYSTEM

[Monji ninshikigo kyori hoshiki]

Inventors:	Hideaki Sugahara and Eiichiro Yamamoto
Applicant:	Denshikeisanki Kihongijutsu Kenkyu Kumiai [Computer Basic Technology Research Consortium]

[There are no amendments to this patent.]

Claim

A post-character-recognition processing system characterized by the following facts: the post-character-recognition processing system has a character recognition means that recognizes read characters and a matching means that detects agreement between said recognized character and the character contained in a word dictionary; in this post-character-recognition processing system, recognition character candidates in plural precedence orders are output from the character recognition means and are compared with the words kept in the word dictionary by means of said matching means; the word with the highest matching degree is selected.

Detailed explanation of the invention

The present invention pertains to a post-character-recognition processing system. In particular, this invention pertains to a post-character-recognition processing system characterized by the fact that the input character input with a character read means is subject to character recognition processing by means of a character dictionary (such as a kanji dictionary), and, after the character recognition processing, on the basis of the recognition results of plural precedence orders, matching is performed with a word dictionary so that the input word can be recognized correctly.

In the conventional character recognition system, as shown in Figure 1, characteristics of the input characters are extracted with recognition unit (1), and the extracted characteristics are compared with a file, and the characters with a high recognition precedence order are output to output register (2); then, as a post-character-recognition processing, if it is known beforehand that the three characters output to said output register (2) indicate a name of a prefecture, these characters are sequentially compared with prefecture dictionary (3) by means of matching circuit (4), so that the input characters are recognized correctly.

As shown in Figure 1, a data input paper (not shown in the figure) with three characters written in the region for writing the name of a prefecture is read with an OCR (not shown in the figure). Based on the obtained data, in recognition unit (1), characters "MIYA", "SAKI", and "PREFECTURE" having the highest recognition precedence order are output. Then, in matching circuit (4), they are sequentially compared with the names of prefectures listed in prefecture dictionary (3), and the name of the prefecture with the highest degree of agreement is read and output. In this post-processing system, as shown in Figure 1, when "MIYA", "SAKI", "PREFECTURE" are output from recognition unit (1) and are matched with the names of prefectures listed in prefecture dictionary (3), there are two names with the same precedence order, namely, "MIYASAKI PREFECTURE" and "MIYAGI PREFECTURE", and it is impossible to select one of them automatically. In this case, an operator is needed to perform the recognition operation control. As a result, it takes a long time in performing the post-processing. This is undesirable.

The purpose of this invention is to solve the aforementioned problems of the conventional methods by providing a post-character-recognition processing system characterized by the following facts: output from the recognition unit is not limited only to one having the highest precedence order; instead, a plurality, say, the first three of them or the first five of them with the highest precedence order, are output; for these plural recognition outputs, matching with the word dictionary is carried out. That is, this invention provides a post-character-recognition processing system characterized by the following facts: the post-character-recognition processing system has a character recognition means that recognizes read characters and a matching means

that detects agreement between said recognized character and the character contained in a word dictionary; in this post-character-recognition processing system, recognition character candidates in plural precedence orders are output from the character recognition means and are compared with the words kept in the word dictionary by means of said matching means; the word with the highest matching degree is selected.

Before describing this invention in detail, the operating principle will be explained briefly with reference to Figure 2.

When the recognition unit reads a 3-character prefecture name, as shown in Figure 2, for the first character, the first precedence order is "KA", the second precedence order is "CHITSU", the third precedence order is "AKI", the fourth precedence order is "ZAI", and the fifth precedence order is "HAYASHI". For the second character, the first precedence order is "TA", the second precedence order is "UCHI", the third precedence order is "KUCHI", the fourth precedence order is "EN", and the fifth precedence order is "YU". For the third character, the first precedence order is "GU", the second precedence order is "KEN", the third precedence order is "ME", the fourth precedence order is "SHA", and the fifth precedence order is "TAN". In this way, plural candidates are obtained in said recognition precedence orders. Then, these candidate characters are sequentially compared with the prefecture names contained in the prefecture dictionary that stores all of the names of the prefectures in Japan. That is, from prefecture dictionary (3), "HOK KAI DO" is first read out. Its first character "HOK" is compared with said "KA, CHITSU, AKI, ZAI, HAYASHI" seeking matching. In this case, it is found that "HOK" is not in agreement with "KA, CHITSU, AKI, ZAI, HAYASHI". Then, the second character, namely, "KAI" is compared with said "TA, UCHI, KUCHI, EN, YU", and the third character "DO" is compared with said "GU, KEN, ME, SHA, TAN" seeking matching. In this case, as there is no fourth character, only the first word "HOK KAI DO" in prefecture dictionary (3) is in agreement with the output of the recognition unit.

However, for the second word in prefecture dictionary (3), namely, "AO MORI KEN", its third character "KEN" is in agreement with one of said "GU, KEN, ME, SHA, TAN", and it also has no four character. This is also a point of agreement. Consequently, the matching degree of the second word "AO MORI KEN" is even higher than the first word "HOK KAI DO".

For the third word "AKI TA KEN" in word dictionary (3), all of the characters are in agreement with said candidate characters. That is, "AKI" is in agreement with one of "KA, CHITSU, AKI, ZAI, HAYASHI"; "TA" is in agreement with one of "TA, UCHI, KUCHI, EN, YU"; and "KEN" is in agreement with one of "GU, KEN, ME, SHA, TAN". Also, it does not have a fourth character. This is also a point of agreement. Consequently, in this case, the matching degree of the third word is the highest. Consequently, the third word "AKI TA KEN" is output as the recognized read characters.

To illustrate advantage, an application example of this invention will be explained with reference to Figure 3.

In this figure, the same part numbers as aforementioned are adopted. (5) represents a character matrix register; (6) represents a precedence order register; (7) represents a word dictionary; (8) represents a matching result output register; (9) represents a result judging circuit; (10) represents an output register; and (11) represents a matching circuit that corresponds to matching circuit (4).

Character matrix register (5) is a register that sets plural candidate characters output from recognition unit (1). For example, for the first character, it sets first through fifth characters of "KA, CHITSU, AKI, ZAI, HAYASHI" as recognition precedence orders [1]-[5]. For the second and third characters, too, it sets "TA, UCHI, KUCHI, EN, YU" and "GU, KEN, ME, SHA, TAN" as the first through fifth precedence orders, respectively.

Precedence order register (6) is a register for output of characters set in character matrix register (5). It sequentially changes the set characters under sequence control signal C_1.

Word dictionary (7) is a file of plural word collections classified as word collection for prefecture names and names of cities of each prefecture. It is classified under control signal C_2 for each word from matching circuit (11), and sequentially outputs the words of the prescribed class in a prescribed sequence.

Matching result output register (8) is a register that holds the degree of agreement between characters set in character matrix register (5) and words output from word dictionary (7) corresponding to words.

Result judging circuit (9) selects and outputs the word with the highest matching degree as a result of matching performed in matching circuit (11).

In the following, we will examine the operation in Figure 3.

(1) The recognition candidate characters output from recognition unit (1) are output to character matrix register (5) in the recognition precedence order. For example, as shown in Figure 2, for the first character, the first through fifth precedence orders of "KA, CHITSU, AKI, ZAI, HAYASHI" are output; for the second character, "TA, UCHI, KUCHI, EN, YU" are output; and, for the third character, "GU, KEN, ME, SHA, TAN" are output. Because it is known beforehand that the output of said recognition unit (1) is a name of a prefecture, the word collection file portion is read out from word dictionary (7). In this case, first of all, "HOK KAI DO" is read out under control signal C_2 of each word output from matching circuit (11). Then, "KA TA GU" are set in precedence order register (6) under precedence order control signal C_1 from matching circuit (11), and are compared with "HOK KAI DO". In this case, as there is no fourth character, this is a point of agreement, while there is no other agreement. Then, from matching circuit (11), precedence order control signal C_1 is output, "CHITSU UCHI KEN" are

set in precedence order register (6) and are compared with "HOK KAI DO" in the same way. In this way, under precedence order control signal C_1 from matching circuit (11), "AKI KUCHI ME" as the third precedence order, "ZAI EN SHA" as the fourth precedence order, and "HAYASHI YU TAN" as the fifth precedence order are sequentially set in precedence order register (6), and are matched with "HOK KAI DO". As these characters are all not in agreement with each other, and only the fact that the fourth character is absent is in agreement, "1" is filled in (4) of class 1 of matching result output register (8), and "0" is filled in (1)-(3) of said class 1.

(2) In this way, as comparison with the first word "HOK KAI DO" comes to an end, matching circuit (11) outputs control signal C_2, and the second word "AO MORI KEN" is output. Then, precedence order control signal C_1 is output, and first precedence order "KA TA GU" ~ fifth precedence order "HAYASHI YU TAN" are sequentially set in precedence order register (6), and compared with said "AO MORI KEN". In this case, for the second precedence order "CHITSU UCHI KEN", agreement takes place with respect to two points, that is, "KEN" and the fact that there is no fourth character. Consequently, matching circuit (11) fills "1" in (3) and (4) of class 2 of matching result output register (8), and "0" in (1), (2) of class (2).

(3) Then, under control signal C_2, matching circuit (11) outputs the third word "AKI TA KEN". Then, in the same way as in said (1) and (2), "KA TA GU" ~ "HAYASHI YU TAN" are sequentially set in precedence order register (6), and are compared with said "AKI TA KEN". In this case, agreement is reached with respect to "TA" in the first precedence order "KA TA GU", "KEN" in the second precedence order "CHITSU UCHI KEN", "AKI" in the third precedence order "AKI KUCHI ME", and the fact that there is no fourth character. Consequently, "1" is filled in (1)-(4) of class (3) of matching result output register (8).

(4) In this way, when all of the names of prefectures have been compared, based on the contents of the various classes of matching result output register (8), result judging circuit (9) finally reads and outputs the third prefecture name "AKI TA KEN" from the results of comparison of class (3) with the highest matching degree. In this way, one can correctly determine "AKI TA KEN" in the post-processing.

Also, as shown in Figure 4, in result judging circuit (9'), first input register (12), second input register (13), and comparison control unit (14) are set. From matching circuit (11), various matching states are input to first input register (12), and are compared with those input beforehand and kept in second input register (13) to determine the matching state. When the matching degree is high for first input register (12) that is newly transmitted, it is loaded in second input register (13). On the other hand, when the matching degree is low, the matching degree for the next word is input as it is to first input register (12). In this constitution, matching result output register (8) shown in Figure 3 becomes unnecessary, and even when there are many words for comparison read from the word dictionary, a simple structure can handle it.

In the above, an explanation has been made for an example of selection of the candidate characters until the fifth precedence order. However, this invention is not limited to this scheme. For example, as shown in Figure 5, changes can be made in this respect. In the example shown in Figure 5, the first precedence order is "MIYA SAKI KEN". Also, the second precedence order is "KAN SAKI GU", and the third precedence order is "TOMI TOGE TAN". They are also compared with prefecture dictionary (3) for matching. As a result, it is possible to correctly determine that the input characters are "MIYA SAKI KEN". Consequently, the problem in the prior art shown in Figure 1 can be solved correctly until the third precedence order.

Brief description of the figures

Figure 1 is a diagram illustrating the conventional post-character-recognition processing system. Figure 2 is a diagram illustrating the principle of operation of this invention. Figure 3 is a diagram illustrating the constitution in an application example of this invention. Figure 4 is a diagram illustrating another application example of the result judging circuit in said application example. Figure 5 is a diagram illustrating the case when the candidate characters are up to the third precedence order.

1	Recognition unit
2	Output register
3	Prefecture dictionary
4	Matching circuit
5, 5'	Character matching register
6	Precedence order register
7	Word dictionary
8	Matching result output register
9, 9'	Result judging circuit
10	Output register
11	Matching circuit
12	First input register
13	Second input register
14	Comparison control unit

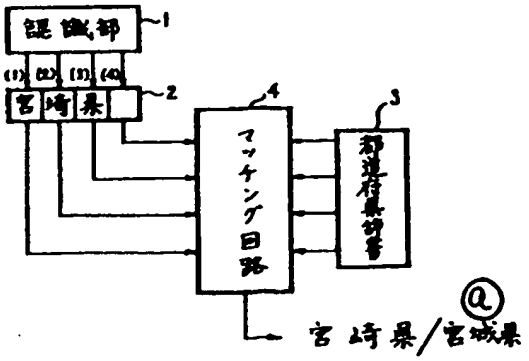


Figure 1

Key: a MIYA SAKI KEN / MIYA GI KEN
1 Recognition unit
2
(1) MIYA
(2) SAKI
(3) KEN
3 Prefecture dictionary
4 Matching circuit

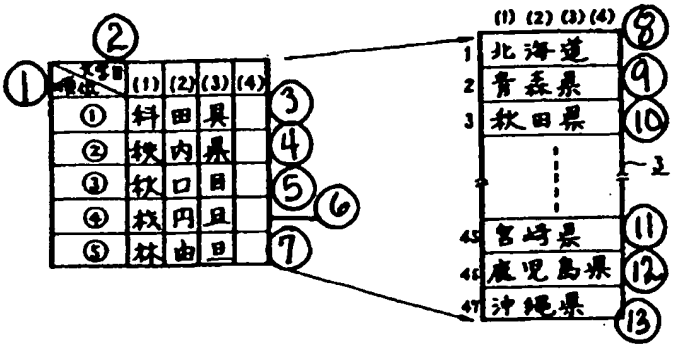


Figure 2

Key: 1 Precedence order
2 Character
3 KA TA GU
4 CHITSU UCHI KEN
5 AKI KUCHI ME
6 KI EN TAN
7 HAYASHI YU TAN
8 HO KAI DO
9 AO MORI KEN
10 AKI TA KEN

11 MIYA SAKI KEN
12 KA GO SHIMA KEN
13 OKI NAWA KEN

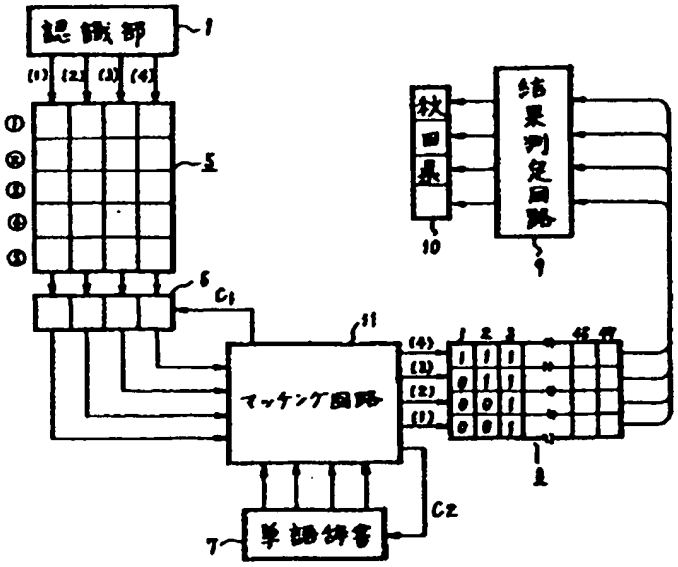


Figure 3

Key: 1 Recognition unit
7 Word dictionary
9 Result judging circuit
10 AKI TA KEN
11 Matching circuit
14 Comparison control

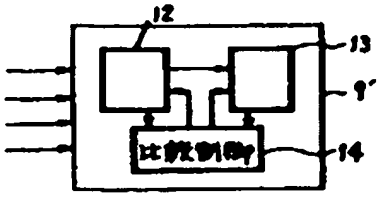


Figure 4

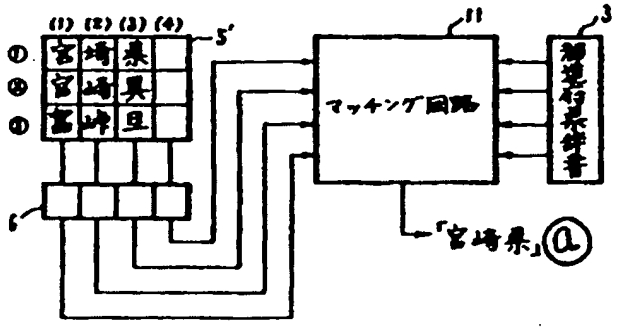


Figure 5

Key: a MIYA SAKI KEN
3 Prefecture dictionary
5'
[1] MIYA SAKI KEN
[2] MIYA SAKI GU
[3] TOMI TOGE TAN
11 Matching circuit

⑲ 日本国特許庁 (JP)　　⑪ 特許出願公開

⑫ 公開特許公報 (A)　　昭58—56189

⑤Int. Cl.³	識別記号	庁内整理番号	⑬公開　昭和58年(1983) 4 月 2 日
G 06 K 9/62		7323—5B	
9/68		7323—5B	発明の数　1
			審査請求　未請求

(全 5 頁)

⑤④文字認識後処理方式

⑳特　　願　昭56—155629
㉒出　　願　昭56(1981) 9 月30日

㉑発 明 者　菅原秀明
川崎市中原区上小田中1015番地
富士通株式会社内

㉑発 明 者　山本栄一郎
川崎市中原区上小田中1015番地
富士通株式会社内

㉑出 願 人　電子計算機基本技術研究組合
東京都港区三田 1 丁目 4 番28号

㉔代 理 人　弁理士　山谷晧栄

明　細　書

1. 発明の名称　文字認識後処理方式

2. 特許請求の範囲

(1) 読取文字を認識する文字認識手段と該認識した文字が単語辞書に記入された文字と一致することを検出するマッチング手段を具備する文字認識後処理方式において、文字認識手段より複数順位の認識文字候補を出力させかつマッチング手段において単語辞書に保持された単語と前記複数順位の認識文字候補と比較してそのマッチング度合のもつとも大きい単語を選択するようにしたことを特徴とする文字認識後処理方式。

3. 発明の詳細な説明

本発明は文字認識後処理方式に関するものであつて、特に文字読取手段により入力された入力文字を文字辞書（例えば漢字辞書）と文字認識処理後の複数順位の認識結果にもとづきさらに単語辞書とのマッチングを行なうことにより入力単語を正確に認識できるようにした文字認識後処理方式に関するものである。

従来の文字認識方式では、例えば第1図に示す如く、認識部1において入力文字の特徴抽出を行ないこれをファイルと比較してもつとも認識順位の高いものを出力レジスタ2に出力し、その後、文字認識後処理としてこの出力レジスタ2に出力された3ケの文字が都道府県名を示すものであるとあらかじめわかつている場合には、これらの文字を都道府県辞書3と順次マッチング回路4にて比較を行ない入力文字を正確に認識するようにしている。

すなわち第1図において都道府県名の記入領域に3個の文字の記入されたデータ入力用紙（図示省略）を例えばＯＣＲ（図示省略）で読取り、これにより得られたデータにもとづき認識部1ではそれぞれ認識順位のもつとも高い「宮」、「埼」「県」を出力レジスタ2に出力し、これらをマッチング回路4において都道府県辞書3にセットさ

れている都道府県名と順次比較してその一致度のもつとも高い都道府県名を読取出力として出力するものである。しかるにこのような後処理方式のものでは、第1図に示す如く、認識部1から「宮」「埼」,「県」と出力されたことにもとづき都道府県辞書3にセットされている都道府県名とマッチングを行なつた場合、「宮崎県」と「宮城県」の2つが同一優先順位で存在することになり、自動的にこのいずれか一方を選択することはできなかつた。そのため場合によつてはオペレータによる認識操作制御を行なう必要が存在するため、その後処理にかなりの時間を必要とする問題が存在する。

したがつて本発明はこのような問題を改善するために認識部からの出力を最高順位のもの1つのみに限定せずに、例えば3番目までのものとかあるいはまた5番目までのものというように、優先順位の高いものから複数個出力させ、これらの複数個の認識出力により単語辞書とのマッチングを行なうようにした文字認識後処理方式を提供することを目的とするものである。そしてこのために本発明における文字認識後処理方式では、読取文字を認識する文字認識手段と該認識した文字が単語辞書に記入された文字と一致することを検出するマッチング手段を具備する文字認識後処理方式において、文字認識手段より複数順位の認識文字候補を出力させかつマッチング手段において単語辞書に保持された単語と前記複数順位の認識文字候補と比較してそのマッチング度合のもつとも大きい単語を選択するようにしたことを特徴とする。

本発明を詳述するに先立ち、第2図にもとづきその動作原理について簡単に説明する。

いま認識部で3文字の都道府県名を読取つたとき第1番目の文字については第2図に示す如く、第1順位が「科」、第2順位が「秩」、第3順位が「秋」、第4順位が「材」、第5順位が「林」であり、第2番目の文字については第1順位が「田」、第2順位が「内」、第3順位が「口」、第4順位が「円」、第5順位が「由」であり、第3番目の文字については第1順位が「具」、第2順位が「県」、第3順位が「目」、第4順位が「且」、第5順位が「旦」とそれぞれ認識順位にしたがつて複数の候補文字が得られたとき、これらの各候補文字を都道府県名の格納された都道府県辞書の各都道府県名と順次比較する。すなわち都道府県辞書3から第1番目に「北海道」を読出し、その第1番目の文字「北」を前記「科，秩，秋，材，林」と比較してマッチングをとる。この場合には「北」と「科，秩，秋，材，林」は不一致である。そして第2番目の文字「海」を前記「田，内，口，円，由」と比較し、第3番目の文字「道」を前記「具，県，目，且，旦」と比較しそれぞれ一致をとる。そしてこの場合には第4番目の文字が存在しないということで認識部の出力と都道府県辞書3の第1番の単語「北海道」とは一致するのみである。

ところが都道府県辞書3の第2番目の単語「青森県」は、その第3番目の文字「県」が前記「具，県，目，且，旦」のうちの1つと一致し、またお互に第4番目の文字が存在しないということでも一致するので、第2番目の単語「青森県」は第1番目の単語「北海道」よりもマッチング度合が大きい。

そして都道府県辞書3の第3番目の単語「秋田県」は、各文字とも前記候補文字と一致する。すなわち「秋」は「科，秩，秋，材，林」の1つに一致し、「田」は「田，内，口，円，由」の1つに一致し、「県」は「具，県，目，且，旦」の1つに一致し、しかも第4番目の文字が存在しないことでも一致する。それ故、この場合、この第3番目の単語とのマッチング度合がもつとも大きいので、この第3番の単語である「秋田県」を読取文字として認識出力するものである。

以下本発明の一実施例を第3図にもとづき説明する。

図中、他図と同符号部は同一部分を示し、5は文字マトリクス・レジスタ、6は順位レジスタ、7は単語辞書、8はマッチング結果出力レジスタ、9は結果判定回路、10は出力レジスタ、11はマッチング回路であつてマッチング回路4に対応

するものである。

文字マトリクス・レジスタ5は認識部1から出力された複数の候補文字がセットされるレジスタであつて、例えば第1番目の文字に対しては第1順位～第5順位までの「科」,「秩」,「秋」,「材」,「林」がその認識順位①～⑤にしたがつてセットされ、第2番目の文字および第3番目の文字に対しても、同様に第1順位～第5順位までの「田」,「内」,「口」,「円」,「由」及び「具」,「県」,「目」,「且」,「且」がセットされる。

順位レジスタ6は文字マトリクス・レジスタ5にセットされている文字が出力されるレジスタであつて、マッチング回路11からの順序制御信号C_1によりそのセットされる文字が順次変更されるものである。

単語辞書7は後処理に必要な、例えば都道府県名用の単語集とか、各都道府県毎の例えば秋田県内の郡市町村名のような分類された複数の単語集がファイルされているものであり、マッチング回路11からの単語毎の制御信号C_2により分類別に、しかも一定の順序にしたがつて所望の分類の単語が順次出力されるものである。

マッチング結果出力レジスタ8は文字マトリクス・レジスタ5にセットされた文字と単語辞書7から出力された単語との一致度を単語対応に保持するレジスタである。

結果判定回路9はマッチング回路11にて行なわれたマッチングの結果、そのもつともマッチング度合の大きな単語を選択出力するものである。

次に第3図の動作について説明する。

(1) 認識部1から出力された認識候補文字はその認識順位にしたがつて文字マトリクス・レジスタ5に出力される。例えば、第2図に示す如く、第1番目の文字に対しては第1順位～第5順位までの「科，秩，秋，材，林」が出力され、第2番目の文字に対しては「田，内，口，円，由」が出力され、第3番目の文字に対しては「具，県，目，且，且」が出力される。そして前記認識部1の出力が都道府県名であることがあらかじめわかつているので、単語辞書7から都道府県名用の単語集ファイル部が読出される。この場合、マッチング回路11から出力される単語毎の制御信号C_2により先ず「北海道」が読出される。そしてマッチング回路11からの順序制御信号C_1により順位レジスタ6に先ず「科田具」がセットされ、「北海道」と比較されるが、このとき4番目に文字がないということのみで一致するが他は一致しない。次にマッチング回路11から再び順序制御信号C_1が出力され、順位レジスタ6に「秩内県」がセットされ同様に「北海道」と比較される。このようにしてマッチング回路11からの順序制御信号C_1により順位レジスタ6に第3順位の「秋口目」、第4順位の「材円且」、第5順位の「林由且」が順次セットされ「北海道」とのマッチングが行なわれるが、これらはいずれも文字同志不一致であり、その結果第4番目の文字が存在しないということで一致するので、マッチング結果出力レジスタ8の区分1の(4)に「1」が記入され区分1の(1)～(3)には「0」が記入される。

(2) このようにして第1番目の単語「北海道」との照合が終るとマッチング回路11は制御信号C_2を出力し、第2番目の単語「青森県」を出力させる。それから順序制御信号C_1を出力し順位レジスタ6に第1順位「科田具」～第5順位「林由且」を順次セットして前記「青森県」と照合する。このとき第2順位の「秩内県」における「県」と第4番目の文字がないという2つの点で一致するので、マッチング回路11はマッチング結果出力レジスタ8の区分2の(3)(4)にそれぞれ「1」が記入され区分2の(1)(2)にそれぞれ「0」が記入される。

(3) 次いでマッチング回路11は制御信号C_2により第3番の単語「秋田県」を出力させ、それから前記(1),(2)と同様にして順位レジスタ6に「科田具」～「林由且」を順次セットしてこの「秋田県」との照合を行なう。この場合には第1順位の「科田具」における「田」、第2順位の「秩内県」の「県」、第3順位の「秋口目」における「秋」と第4番目の文字がないということでそれぞ

れ一致が得られるので、マッチング結果出力レジスタ8の区分3の(1)～(4)にはそれぞれ「1」が記入されることになる。

(4) このようにしてすべての都道府県名との照合が終了したとき、結果判定回路9はこのマッチング結果出力レジスタ8の各区分の内容にもとづきそのもつともマッチング度合の大きな区分3の照合結果により第3番目の都道府県名の「秋田県」を最終的な読取出力として出力レジスタ10に出力する。このようにして後処理により「秋田県」を正確に取出すことができる。

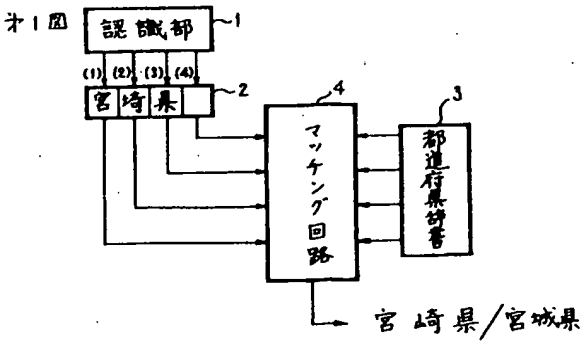
なお第4図に示す如く、結果判定回路9'に第1入力レジスタ12、第2入力レジスタ13および比較制御部14を設け、第1入力レジスタ12にマッチング回路11からの個々のマッチング状態を入力してこれを先に入力している第2入力レジスタ13に保持されているものとのマッチング状態と比較して、新らしく伝達された第1入力レジスタ12のマッチング度が大きいときにこれを第2入力レジスタ13に記入し、小さいときにはそのまま第1入力レジスタ12に次の単語に対するマッチング度を入力するように構成すれば、第3図におけるマッチング結果出力レジスタ8は不必要となり、単語辞書から読出される被照合単語数が大きい場合でも簡単な構成で対処することができる。

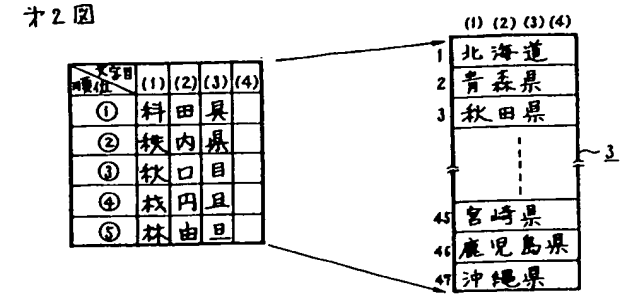
なお上記説明では候補文字を第5順位まで選択した例について説明したが、これに限定されるものではない。例えば第5図に示す如く、これを変更することもできる。第5図の例では、第1順位が「宮埼県」であつても、その第2順位の「宮崎具」、第3順位の「富峠且」を都道府県辞書3と比較してマッチングをとることにより、入力文字が「宮崎県」であるものと正しく識別できるので、第3順位までででも第1図に示した従来のような問題を正しく解決することができる。

4.図面の簡単な説明

第1図は従来の文字認識後処理方式説明図、第2図は本発明の一実施例構成図、第4図はその結果判定回路の他の実施例、第5図は候補文字を第3順位までにした場合の説明図である。

図中、1は認識部、2は出力レジスタ、3は都道府県辞書、4はマッチング回路、5，5'は文字マトリクス・レジスタ、6は順位レジスタ、7は単語辞書、8はマッチング結果出力レジスタ、9，9'は結果判定回路、10は出力レジスタ、11はマッチング回路、12は第1入力レジスタ、13は第2入力レジスタ、14は比較制御部をそれぞれ示す。

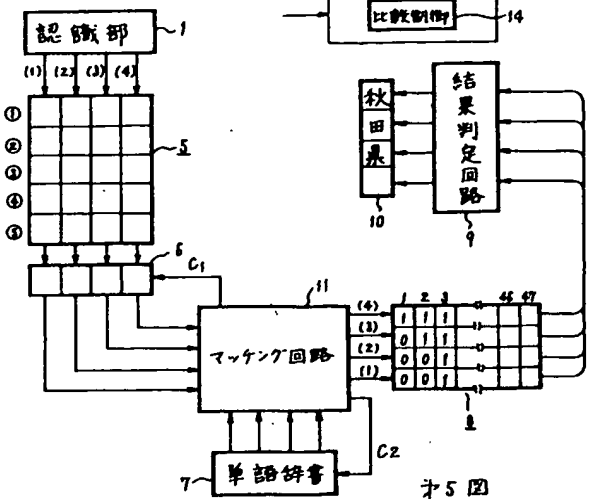




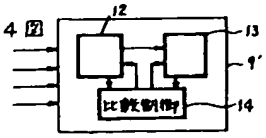
特許出願人　電子計算機基本技術研究組合

代理人弁理士　山谷晴榮

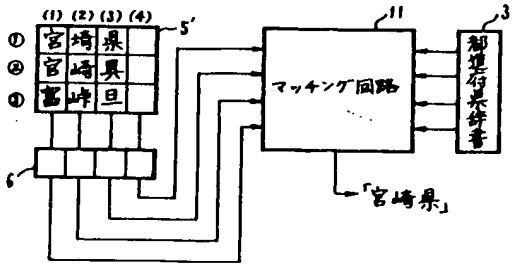
第3図


認識部
1
(1) (2) (3) (4)
①
②
③
④
⑤
5
6
C1
秋
田
県
10
結果判定回路
9
11
マッチング回路
(4)
(3)
(2)
(1)
1 2 3 46 47
1 1 1
0 1 1
0 0 1
0 0 1
8
C2
7
単語辞書

第4図


12
13
9'
比較制御
14

第5図


(1) (2) (3) (4)
①
②
③
宮 埼 県
宮 崎 県
高 岸 旦
5'
6
11
マッチング回路
3
「宮崎県」

PTO 03-567

Japanese Kokai Patent Application
No. Sho 58[1983]-56189

POST-CHARACTER-RECOGNITION PROCESSING SYSTEM

Hideaki Sugahara and Eiichiro Yamamoto

UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. NOVEMBER 2002
TRANSLATED BY THE RALPH MCELROY TRANSLATION COMPANY

JAPANESE PATENT OFFICE
PATENT JOURNAL (A)
KOKAI PATENT APPLICATION NO. SHO 58[1983]-56189

Int. Cl.[3]:	G 06 K 9/62 9/68
Sequence Nos. for Office Use:	7323-5B
Filing No.:	Sho 56[1981]-155629
Filing Date:	September 30, 1981
Publication Date:	April 2, 1983
No. of Inventions:	1 (Total of 5 pages)
Examination Request:	Not filed

POST-CHARACTER-RECOGNITION PROCESSING SYSTEM

[Monji ninshikigo kyori hoshiki]

Inventors:	Hideaki Sugahara and Eiichiro Yamamoto
Applicant:	Denshikeisanki Kihongijutsu Kenkyu Kumiai [Computer Basic Technology Research Consortium]

[There are no amendments to this patent.]

Claim

A post-character-recognition processing system characterized by the following facts: the post-character-recognition processing system has a character recognition means that recognizes read characters and a matching means that detects agreement between said recognized character and the character contained in a word dictionary; in this post-character-recognition processing system, recognition character candidates in plural precedence orders are output from the character recognition means and are compared with the words kept in the word dictionary by means of said matching means; the word with the highest matching degree is selected.

Detailed explanation of the invention

The present invention pertains to a post-character-recognition processing system. In particular, this invention pertains to a post-character-recognition processing system characterized by the fact that the input character input with a character read means is subject to character recognition processing by means of a character dictionary (such as a kanji dictionary), and, after the character recognition processing, on the basis of the recognition results of plural precedence orders, matching is performed with a word dictionary so that the input word can be recognized correctly.

In the conventional character recognition system, as shown in Figure 1, characteristics of the input characters are extracted with recognition unit (1), and the extracted characteristics are compared with a file, and the characters with a high recognition precedence order are output to output register (2); then, as a post-character-recognition processing, if it is known beforehand that the three characters output to said output register (2) indicate a name of a prefecture, these characters are sequentially compared with prefecture dictionary (3) by means of matching circuit (4), so that the input characters are recognized correctly.

As shown in Figure 1, a data input paper (not shown in the figure) with three characters written in the region for writing the name of a prefecture is read with an OCR (not shown in the figure). Based on the obtained data, in recognition unit (1), characters "MIYA", "SAKI", and "PREFECTURE" having the highest recognition precedence order are output. Then, in matching circuit (4), they are sequentially compared with the names of prefectures listed in prefecture dictionary (3), and the name of the prefecture with the highest degree of agreement is read and output. In this post-processing system, as shown in Figure 1, when "MIYA", "SAKI", "PREFECTURE" are output from recognition unit (1) and are matched with the names of prefectures listed in prefecture dictionary (3), there are two names with the same precedence order, namely, "MIYASAKI PREFECTURE" and "MIYAGI PREFECTURE", and it is impossible to select one of them automatically. In this case, an operator is needed to perform the recognition operation control. As a result, it takes a long time in performing the post-processing. This is undesirable.

The purpose of this invention is to solve the aforementioned problems of the conventional methods by providing a post-character-recognition processing system characterized by the following facts: output from the recognition unit is not limited only to one having the highest precedence order; instead, a plurality, say, the first three of them or the first five of them with the highest precedence order, are output; for these plural recognition outputs, matching with the word dictionary is carried out. That is, this invention provides a post-character-recognition processing system characterized by the following facts: the post-character-recognition processing system has a character recognition means that recognizes read characters and a matching means

that detects agreement between said recognized character and the character contained in a word dictionary; in this post-character-recognition processing system, recognition character candidates in plural precedence orders are output from the character recognition means and are compared with the words kept in the word dictionary by means of said matching means; the word with the highest matching degree is selected.

Before describing this invention in detail, the operating principle will be explained briefly with reference to Figure 2.

When the recognition unit reads a 3-character prefecture name, as shown in Figure 2, for the first character, the first precedence order is "KA", the second precedence order is "CHITSU", the third precedence order is "AKI", the fourth precedence order is "ZAI", and the fifth precedence order is "HAYASHI". For the second character, the first precedence order is "TA", the second precedence order is "UCHI", the third precedence order is "KUCHI", the fourth precedence order is "EN", and the fifth precedence order is "YU". For the third character, the first precedence order is "GU", the second precedence order is "KEN", the third precedence order is "ME", the fourth precedence order is "SHA", and the fifth precedence order is "TAN". In this way, plural candidates are obtained in said recognition precedence orders. Then, these candidate characters are sequentially compared with the prefecture names contained in the prefecture dictionary that stores all of the names of the prefectures in Japan. That is, from prefecture dictionary (3), "HOK KAI DO" is first read out. Its first character "HOK" is compared with said "KA, CHITSU, AKI, ZAI, HAYASHI" seeking matching. In this case, it is found that "HOK" is not in agreement with "KA, CHITSU, AKI, ZAI, HAYASHI". Then, the second character, namely, "KAI" is compared with said "TA, UCHI, KUCHI, EN, YU", and the third character "DO" is compared with said "GU, KEN, ME, SHA, TAN" seeking matching. In this case, as there is no fourth character, only the first word "HOK KAI DO" in prefecture dictionary (3) is in agreement with the output of the recognition unit.

However, for the second word in prefecture dictionary (3), namely, "AO MORI KEN", its third character "KEN" is in agreement with one of said "GU, KEN, ME, SHA, TAN", and it also has no four character. This is also a point of agreement. Consequently, the matching degree of the second word "AO MORI KEN" is even higher than the first word "HOK KAI DO".

For the third word "AKI TA KEN" in word dictionary (3), all of the characters are in agreement with said candidate characters. That is, "AKI" is in agreement with one of "KA, CHITSU, AKI, ZAI, HAYASHI"; "TA" is in agreement with one of "TA, UCHI, KUCHI, EN, YU"; and "KEN" is in agreement with one of "GU, KEN, ME, SHA, TAN". Also, it does not have a fourth character. This is also a point of agreement. Consequently, in this case, the matching degree of the third word is the highest. Consequently, the third word "AKI TA KEN" is output as the recognized read characters.

To illustrate advantage, an application example of this invention will be explained with reference to Figure 3.

In this figure, the same part numbers as aforementioned are adopted. (5) represents a character matrix register; (6) represents a precedence order register; (7) represents a word dictionary; (8) represents a matching result output register; (9) represents a result judging circuit; (10) represents an output register; and (11) represents a matching circuit that corresponds to matching circuit (4).

Character matrix register (5) is a register that sets plural candidate characters output from recognition unit (1). For example, for the first character, it sets first through fifth characters of "KA, CHITSU, AKI, ZAI, HAYASHI" as recognition precedence orders [1]-[5]. For the second and third characters, too, it sets "TA, UCHI, KUCHI, EN, YU" and "GU, KEN, ME, SHA, TAN" as the first through fifth precedence orders, respectively.

Precedence order register (6) is a register for output of characters set in character matrix register (5). It sequentially changes the set characters under sequence control signal C_1.

Word dictionary (7) is a file of plural word collections classified as word collection for prefecture names and names of cities of each prefecture. It is classified under control signal C_2 for each word from matching circuit (11), and sequentially outputs the words of the prescribed class in a prescribed sequence.

Matching result output register (8) is a register that holds the degree of agreement between characters set in character matrix register (5) and words output from word dictionary (7) corresponding to words.

Result judging circuit (9) selects and outputs the word with the highest matching degree as a result of matching performed in matching circuit (11).

In the following, we will examine the operation in Figure 3.

(1) The recognition candidate characters output from recognition unit (1) are output to character matrix register (5) in the recognition precedence order. For example, as shown in Figure 2, for the first character, the first through fifth precedence orders of "KA, CHITSU, AKI, ZAI, HAYASHI" are output; for the second character, "TA, UCHI, KUCHI, EN, YU" are output; and, for the third character, "GU, KEN, ME, SHA, TAN" are output. Because it is known beforehand that the output of said recognition unit (1) is a name of a prefecture, the word collection file portion is read out from word dictionary (7). In this case, first of all, "HOK KAI DO" is read out under control signal C_2 of each word output from matching circuit (11). Then, "KA TA GU" are set in precedence order register (6) under precedence order control signal C_1 from matching circuit (11), and are compared with "HOK KAI DO". In this case, as there is no fourth character, this is a point of agreement, while there is no other agreement. Then, from matching circuit (11), precedence order control signal C_1 is output, "CHITSU UCHI KEN" are

set in precedence order register (6) and are compared with "HOK KAI DO" in the same way. In this way, under precedence order control signal C_1 from matching circuit (11), "AKI KUCHI ME" as the third precedence order, "ZAI EN SHA" as the fourth precedence order, and "HAYASHI YU TAN" as the fifth precedence order are sequentially set in precedence order register (6), and are matched with "HOK KAI DO". As these characters are all not in agreement with each other, and only the fact that the fourth character is absent is in agreement, "1" is filled in (4) of class 1 of matching result output register (8), and "0" is filled in (1)-(3) of said class 1.

(2) In this way, as comparison with the first word "HOK KAI DO" comes to an end, matching circuit (11) outputs control signal C_2, and the second word "AO MORI KEN" is output. Then, precedence order control signal C_1 is output, and first precedence order "KA TA GU" ~ fifth precedence order "HAYASHI YU TAN" are sequentially set in precedence order register (6), and compared with said "AO MORI KEN". In this case, for the second precedence order "CHITSU UCHI KEN", agreement takes place with respect to two points, that is, "KEN" and the fact that there is no fourth character. Consequently, matching circuit (11) fills "1" in (3) and (4) of class 2 of matching result output register (8), and "0" in (1), (2) of class (2).

(3) Then, under control signal C_2, matching circuit (11) outputs the third word "AKI TA KEN". Then, in the same way as in said (1) and (2), "KA TA GU" ~ "HAYASHI YU TAN" are sequentially set in precedence order register (6), and are compared with said "AKI TA KEN". In this case, agreement is reached with respect to "TA" in the first precedence order "KA TA GU", "KEN" in the second precedence order "CHITSU UCHI KEN", "AKI" in the third precedence order "AKI KUCHI ME", and the fact that there is no fourth character. Consequently, "1" is filled in (1)-(4) of class (3) of matching result output register (8).

(4) In this way, when all of the names of prefectures have been compared, based on the contents of the various classes of matching result output register (8), result judging circuit (9) finally reads and outputs the third prefecture name "AKI TA KEN" from the results of comparison of class (3) with the highest matching degree. In this way, one can correctly determine "AKI TA KEN" in the post-processing.

Also, as shown in Figure 4, in result judging circuit (9'), first input register (12), second input register (13), and comparison control unit (14) are set. From matching circuit (11), various matching states are input to first input register (12), and are compared with those input beforehand and kept in second input register (13) to determine the matching state. When the matching degree is high for first input register (12) that is newly transmitted, it is loaded in second input register (13). On the other hand, when the matching degree is low, the matching degree for the next word is input as it is to first input register (12). In this constitution, matching result output register (8) shown in Figure 3 becomes unnecessary, and even when there are many words for comparison read from the word dictionary, a simple structure can handle it.

In the above, an explanation has been made for an example of selection of the candidate characters until the fifth precedence order. However, this invention is not limited to this scheme. For example, as shown in Figure 5, changes can be made in this respect. In the example shown in Figure 5, the first precedence order is "MIYA SAKI KEN". Also, the second precedence order is "KAN SAKI GU", and the third precedence order is "TOMI TOGE TAN". They are also compared with prefecture dictionary (3) for matching. As a result, it is possible to correctly determine that the input characters are "MIYA SAKI KEN". Consequently, the problem in the prior art shown in Figure 1 can be solved correctly until the third precedence order.

Brief description of the figures

Figure 1 is a diagram illustrating the conventional post-character-recognition processing system. Figure 2 is a diagram illustrating the principle of operation of this invention. Figure 3 is a diagram illustrating the constitution in an application example of this invention. Figure 4 is a diagram illustrating another application example of the result judging circuit in said application example. Figure 5 is a diagram illustrating the case when the candidate characters are up to the third precedence order.

1	Recognition unit
2	Output register
3	Prefecture dictionary
4	Matching circuit
5, 5'	Character matching register
6	Precedence order register
7	Word dictionary
8	Matching result output register
9, 9'	Result judging circuit
10	Output register
11	Matching circuit
12	First input register
13	Second input register
14	Comparison control unit

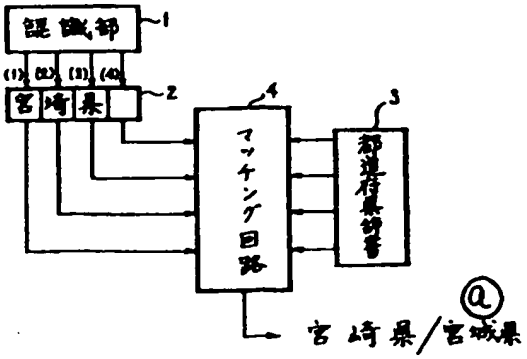


Figure 1

Key: a MIYA SAKI KEN / MIYA GI KEN
1 Recognition unit
2
(1) MIYA
(2) SAKI
(3) KEN
3 Prefecture dictionary
4 Matching circuit

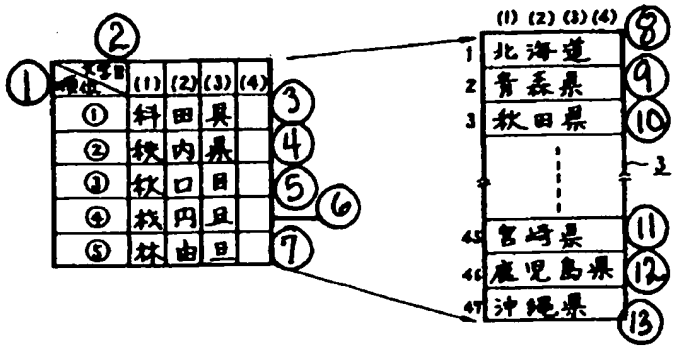


Figure 2

Key: 1 Precedence order
2 Character
3 KA TA GU
4 CHITSU UCHI KEN
5 AKI KUCHI ME
6 KI EN TAN
7 HAYASHI YU TAN
8 HO KAI DO
9 AO MORI KEN
10 AKI TA KEN

11 MIYA SAKI KEN
12 KA GO SHIMA KEN
13 OKI NAWA KEN

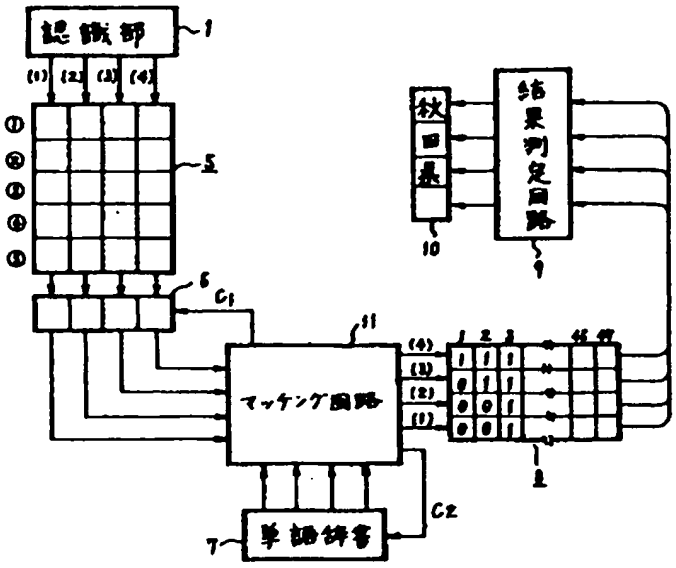


Figure 3

Key: 1 Recognition unit
7 Word dictionary
9 Result judging circuit
10 AKI TA KEN
11 Matching circuit
14 Comparison control

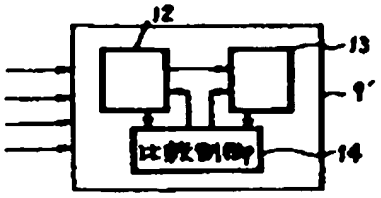


Figure 4

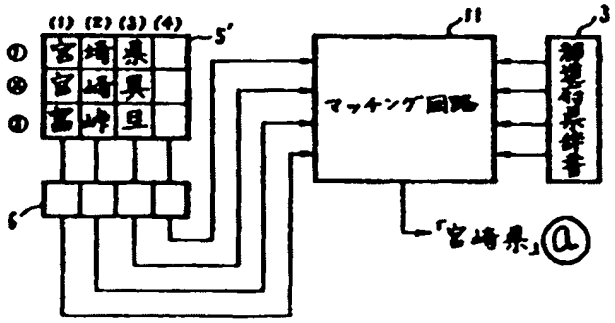


Figure 5

Key: a MIYA SAKI KEN
3 Prefecture dictionary
5'
[1] MIYA SAKI KEN
[2] MIYA SAKI GU
[3] TOMI TOGE TAN
11 Matching circuit